

**The Evolution of Global  
Bilateral Migration 1960-2000**  
by  
**Özden, Parsons, Schiff, Walmsley**

**Discussion by**

**Thomas MELONIO, AFD Research Department**

International conference on Migration and Development  
Washington, September 10, 2009.

*développeur d'avenirs durables*

**afd**  
AGENCE FRANÇAISE  
DE DÉVELOPPEMENT

## Paper content

- The dataset consists of the addition of 5 global matrices of bilateral migrant stocks disaggregated by gender, for the following years 1960, 1970, 1980, 1990, 2000
- All 5 matrices are square, with 226 origin and destination countries
- Main primary source of data is the *Global Migrant database*, built by the UNDP, the UN statistics division, the WB and the University of Sussex

## Paper content

- Migrants are those identified as foreign-born in national censuses or population registers.
- Compared to previous existing databases, this one provides with more depth in both the time dimension (1960-2000) and the geographical dimension (226\*226).
- This database does not originate from any new data collection, it is based on harmonization of existing data and the authors' estimations to fill missing values.

## Main findings

- The global migrants' stock has not increased as a percentage of the world population btw 1960 and 2000 (from 2,5% to 2,6%), even though the global migrant stock doubled, from 76 to 159 million (World Pop.: 3 billion ► 6 billion). Both grew at an average annual rate of 1.75% over that period.
- South-North migration has been the most dynamic flow, its growth in number outpacing that of S-S, N-S and N-N migration during all periods, except S-S migration during the 90'.

## General comments

- The database, in its scope and ambition, somewhat recalls of great datasets such as Barro and Lee's
- This paper is very clear and all data treatments are presented in a very transparent way.
- Out of the 510 000 bilateral migration stocks given in the dataset, it's not very clear what percentage originates from data collection & harmonization and what comes from new estimations. Display that percentage!

# Quick data analysis

Region	Nb of censuses available	Theoretical number of censuses	% available
Aus, NZ, Jap, Can, USA	22	25	88%
Western Europe	107	130	82%
Eastern Europe and central Asia	47	140	34%
MENA	46	95	48%
Africa	137	255	54%
South Asia	20	40	50%
East Asia and the Pacific	124	200	62%
Latin America and the Caribbean	195	245	80%
<b>Total</b>	<b>698</b>	<b>1130</b>	<b>62%</b>

## Quick data analysis

- According to the authors, countries with less than one census missing (out of five) accounted for 53% of the global migrant stock in 2000 (or 65% excluding the former soviet union). That's still 128 countries with two or more censuses missing...
- Should countries with zero or one census should be abandoned (there are 35 such countries)? In these countries migration stocks cannot be estimated using extrapolation or interpolation. In the absence of any measured national immigration trend, is it reliable to “import” a regional trend?

## Country-specific issues

- French Guiana, Guadeloupe, Martinique and St Pierre et Miquelon are not independent states but instead full-right French « regions »: the World Bank does not officially support independence in those territories!!!
- It is mentioned that China and Vietnam have a different definition of migrants than that of nother migrants, is it really impossible to include them?

# Methodological discussion

## **What happens when data is only available at the aggregate level?**

- In many censuses, immigrants are registered by region of origin, not country of origin, which requires treatment to match the structure of the dataset.
- The favored option in this paper is to use data from other decades as the basis for disaggregation. The relative propensity to host migrants from one country is averaged over time.

# Disaggregation issues

## Equation 1: Average destination country shares

$$M_{o,d}^{t,g} = \left( \frac{\sum_{k \neq t} M_{o,d}^{k, "total"}}{\sum_{k \neq t} \sum_{i \in J} M_{i,d}^{k, "total"}} \right) \sum_{i \in J} M_{i,d}^{t,g}$$

$o$  = the country of origin (for example, Armenia)

$d$  = the country of destination (in this case, Australia)

$k$  = all the decades under study i.e. 1960 to 2000 for which there is data (in this case, only 2000)

$t$  = the decade in question (in this case, 1970)

$g$  = is the gender type of migrants, either 'male' or female'

$J$  = the set of all countries in the aggregated category in which  $o$  belongs (in this case, USSR)

$i$  = a single country in the aggregate category  $J$ .

**Table 3. Allocation of Aggregate origin region, according to preferred method**

Origin Country listed in 2001 Australian census	Total immigrants to Australia in 2001	Share of USSR migration to Australia in 2001	Number migrants allocated in 1966 across constituent countries
Azerbaijan	145	0.3%	93
Armenia	899	2.0%	576
Belarus	1,041	2.3%	667
...	...	...	...
Moldova	483	1.1%	309
Russian Federation	15,022	32.8%	9,625
Tajikistan	41	0.1%	26
Turkmenistan	26	0.1%	17
Ukraine	14,062	30.7%	9,010
Uzbekistan	412	0.9%	264
TOTAL USSR	45,745	100%	29,311

# Using averages is suboptimal

Country E (dest.)	1960	1970	1980	1990	2000	Average	Interpolation	Linear pred.
Country A share	10	na	30	40	50	32.5	20	20
Country B share	40	na	30	10	0	20	35	33
Country C share	0	na	0	30	40	17.5	0	4
Country D share	50	na	40	20	10	30	45	43
Total	100	100	100	100	100	100	100	100

# Disaggregation issues

- The slide shown above is an extreme example with rapidly moving migration stocks. Still, in the case of census data, interpolation should be the preferred option when available, linear estimation being more suitable when extrapolation is needed!
- The same stays true when no national data is available and regional immigration propensities have to be used.
- When no regional data is available, does it makes much sense to use global shares? However, it seems that this case is very rare in the dataset.

## Missing data issues

- Apart from disaggregation issues, some censuses are just missing.
- The authors have applied two treatments in that occurrence. Interpolation has been applied when possible (which is ok), when not, it is assumed that the destination country has the same bilateral migrant composition as the decade closest to the missing period. Again, a linear estimation, when possible, would be an improvement.
- In the case of population/migration growth, assuming a geometric progression should be a preferred option (for migration flows and stocks, not migration shares)

## Adjusting for censuses dates

- Because censuses do not occur exactly at the desired dates (1960-70-80-90-2000), the decision has been taken to match each census with its best match, keeping the numbers of that census unchanged. For example, 1962 is matched with 1960, 1979 with 1980, etc.

- Instead of keeping these figures unchanged, it would be an improvement to make a linear estimation (or assume a geometric growth). Example: France has taken censuses in 62, 68, 82, 90 and 99. Instead of saying that 1962=1960, the 1960 numbers could be estimated using the 62-68 trend; 1970 using the 62-82 trend and so on. This would require only a little more programming and bring more comparability across countries.

# Robustness checking

- It would be useful to run robustness tests to convince the reader/user that estimates of disaggregated/missing values are unbiased or have reasonable variance.
- You could flag 2000/3000 « real » cells consisting of real data collection originally disaggregated, re-estimate them through the techniques you used and show the results!

## Using the *global bilateral migration database*

- It may be wiser to drop USSR and India for macroanalysis: both generate large (and as recognized in the paper) somewhat virtual numbers.
- Beware of running statistical regressions of data already produced through linear regressions (or worse techniques). A remark relevant for other datasets such as Barro and Lee's mentioned earlier (see for example Cohen and Soto, 2007).

# Using the global bilateral migration database

- A very good tool for policymakers,
- A huge breakthrough in the estimation of migrants' stocks in developing countries,
- A reliable dataset to extract descriptive data or produce new descriptive data, such as migration balances (Melonio, 2008) for example
- A potential powerful resource to analyze the long term determinants of international migration, with some robustness tests still needed.

**Thank you for your attention**

**Thomas MELONIO**

*développeur d'avenirs durables*

